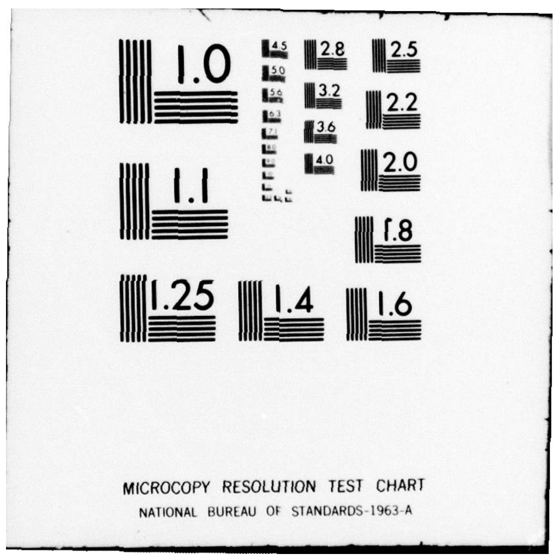
AD-A058 963 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF BIOSTATISTICS F/G 12/1
ON JACKKNIFING IN ESTIMATING THE FINITE END-POINTS OF A DISTRIB--ETC(U)
1978 P K SEN AFOSR-74-2736

UNCLASSIFIED AFOSR-TR-78-1271 NL

1 OF 1
AD
A058963

END
DATE
FILMED
11-78
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(18) AFOSR (19) TR-78-1271

(2) LEVEL II

(6) ON JACKKNIFING IN ESTIMATING THE FINITE END-POINTS OF A DISTRIBUTION*

by

(10) Pranab Kumar Sen
University of North Carolina, Chapel Hill

(9) Interim rept.

ABSTRACT

(11) 1978

(12) 11p.

Sample extreme values are biased estimators of the end-points of a distribution, and hence, jackknifing is useful. However, the properties of jackknifing in such a case differ considerably from those in the regular case. These are studied here. Along with a modification of jackknifing, some applications are also considered.

(16) 2304

AMS 1970 Classification No. 62G30, 62F99.

(17) A5

Key Words & Phrases: bias; extreme values; jackknifing; mean square; order of terminal contact; studentized form; Tukey-estimator of variance.

(15)

*Work supported by the Air Force Office of Scientific Research, A.F.S.C., U.S.A.F., Grant No. AFOSR-74-2736. Reproduction in part or whole permitted for any purpose of the U.S. Government.

See 1473

408 985

78 09 05 007

LB

1. INTRODUCTION

Let $\{X_i, i \geq 1\}$ be a sequence of independent and identically distributed random variables (i.i.d.r.v.) with a distribution function (df) F, defined on $(-\infty,\infty)$. It is assumed that F has a finite (unknown) lower end-point θ, that is

$$(1.1) \qquad -\infty < \theta = \sup\{x: F(x) = 0\} < \infty$$

and that $F(x)$ is continuous and monotonic in $x \in (\theta, \theta + \delta)$, for some $\delta > 0$. A natural estimator of θ is the sample minimum i.e.,

$$(1.2) \qquad \hat{\theta}_n = \min\{X_1,\ldots,X_n\} = X_{n,1} \qquad (n \geq 1) ,$$

where $X_{n,1} \leq \ldots \leq X_{n,n}$ stand for the ordered variables corresponding to $X_1,\ldots,X_n$; $n \geq 1$. $\hat{\theta}_n$ is a (strongly) consistent estimator of θ, but it is not an unbiased one; the nature of its bias depends on the order of terminal contact of F (at θ). It may therefore be appealing to use the jackknife estimator corresponding to $\hat{\theta}_n$.

Under quite general regularity conditions (viz., [1,2,4]), jackknifing meets three objectives: (a) *Bias reduction*. If θ_n^* be the jackknife estimator then $nE(\theta_n^* - \theta) \to 0$ as $n \to \infty$. (b) *Asymptotic normality*. If $n^{1/2}(\hat{\theta}_n - \theta)$ is asymptotically normal, then the same limit law holds for $n^{1/2}(\theta_n^* - \theta)$. (c) The Tukey estimator v_n^2 [defined by (2.5)] is a (strongly) consistent estimator of the variance of $n^{1/2}(\theta_n^* - \theta)$.

Since the asymptotic distributions of sample extrema are non-normal and, depending on the order of terminal contact, the bias of $\hat{\theta}_n$ is $O(n^{-a})$

for some $0 < a \le 1$, the effectiveness of jackknifing in regard to (a) and (b) remains to be examined carefully. Further, in this case, v_n^2 does not converge (stochastically). Along with the preliminary notions, expressions for θ_n^* and v_n^2 are considered in Section 2. The main results are studied in Section 3. Section 4 deals with a modification of jackknifing appropriate for the case of the bias of $O(n^{-a})$ for some $a < 1$. Some general remarks are made in the concluding section.

2. PRELIMINARY NOTIONS

We assume that for some non-negative integer m, $F(x)$ has continuous jth derivative $F^{(j)}(x)[= f^{(j-1)}(x)]$ for all $x \in (\theta, \theta+\delta)$, $\delta > 0$, $1 \le j \le m+1$. We denote the (right hand) derivatives at θ by $F_+^{(j)}(\theta) = f_+^{(j-1)}(\theta)$, $1 \le j \le m+1$ and $F_+^{(0)}(\theta) = 0$, $f_+^{(0)}(\theta) = f_+(\theta)$. Then, a *terminal contact of order* m is defined by

$$(2.1) \qquad F_+^{(j)}(\theta) = 0\ ,\quad 0 \le j \le m \quad \text{and} \quad 0 < f_+^{(m)}(\theta) < \infty\ .$$

Also, for the study of the bias, we assume that

$$(2.2) \qquad \nu_\alpha = \int_\theta^\infty |x|^\alpha dF(x) < \infty \quad \text{for some} \quad \alpha > 0\ .$$

To define θ_n^*, we let for each i: $1 \le i \le n$,

$$(2.3)\ \hat{\theta}_{n-1}^i = \min\{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n\}\ ,\quad \hat{\theta}_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_{n-1}^i\ .$$

Then, $\hat{\theta}_{n-1}^i$ is equal to $X_{n,1}$ or $X_{n,2}$ according as X_i is $\ne$ or $= X_{n,1}$, $1 \le i \le n$. Also,

(2.4) $$\theta_n^* = n^{-1}\sum_{i=1}^n \hat{\theta}_{n,i}$$

$$= X_{n,1} - n^{-1}(n-1)(X_{n,2} - X_{n,1}) \ , \quad n \geq 2 \ .$$

The Tukey estimator v_n^2, defined by

(2.5) $$v_n^2 = \frac{1}{(n-1)} \sum_{i=1}^n (\hat{\theta}_{n,i} - \theta_n^*)^2 = (n-1)\sum_{i=1}^n (\hat{\theta}_{n-1}^i - \theta_n^*)^2 \ ,$$

reduces in our case to

(2.6) $$v_n^2 = (X_{n,2} - X_{n,1})^2 (n-1)(n^2+n-1)/n \quad (\sim \{n(X_{n,2} - X_{n,1})\}^2) \ .$$

For a terminal contact of order $m(\geq 0)$, we define

(2.7) $$b_{n,m} = \{nf_+^{(m)}(\theta)/(m+1)!\}^{1/(m+1)} \ , \quad a_m = 1/(m+1) \ .$$

Then, the limiting distribution of $b_{n,m}(\hat{\theta}_n - \theta)$ is known to be

(2.8) $$\Lambda_m(x) = \begin{cases} 0 \ , & x \leq 0 \ , \\ 1 - \exp(-x^{m+1}) \ , & x > 0 \ . \end{cases}$$

Also, by Theorem 3.1 of Sen (1961), as $n \to \infty$,

(2.9) $$b_{n,m}E(X_{n,r} - \theta) = \Gamma(r+a_m)/\Gamma(r) + o(1) \ , \quad \text{for every (fixed) } r(\geq 1) \ .$$

3. BASIC PROPERTIES OF JACKKNIFING

It follows from (2.4) that

(3.1) $$n(\theta_n^* - \theta) = n(X_{n,1} - \theta) - (n-1)(X_{n,2} - X_{n,1})$$

$$= (2n-1)(X_{n,1} - \theta) - (n-1)(X_{n,2} - \theta) \ .$$

Hence, from (2.9) and (3.1), we obtain that for a terminal contact of order m,

$$(3.2) \qquad b_{n,m}E(\theta_n^* - \theta) = (1-a_m)\sqrt{1+a_m} + o(1)$$
$$= (1-a_m)\{b_{n,m}E(\hat{\theta}_n - \theta)\} + o(1) .$$

For $m = 0$ i.e., $a_m = 1$, the right hand side (rhs) of (3.2) converges to 0, as $n \to \infty$, while for $m \geq 1$ (i.e., $a_m \leq \frac{1}{2}$), jackknifing leads to effectively $100(1-a_m)\%$ reduction in bias. Thus, the basic role of jackknifing is partially impaired for a terminal contact of order $m(\geq 1)$.

<u>Theorem 1</u>. *For a terminal contact of order* $m(\geq 0)$,

$$\Lambda_n^*(x) = \lim_{n\to\infty} P\{b_{n,m}(\theta_n^* - \theta) \leq x\}$$

$$(3.3) \qquad = \begin{cases} \int_0^\infty \exp\left\{-(2y^{a_m} - x)^{m+1}\right\}dy , & -\infty < x \leq 0 , \\ 1 - \exp(-x^{m+1}) + \int_{x^{m+1}}^\infty \exp\left\{-(2y^{a_m} - x)^{m+1}\right\}dy , & x > 0 , \end{cases}$$

where a_m *and* $b_{n,m}$ *are defined by* (2.7).

<u>Proof</u>. Let $Z_n = b_{n,m}(\theta_n^* - \theta)$ and let

$$(3.4) \qquad Y_{n(1)} = nF(X_{n,1}) \quad \text{and} \quad Y_{n(2)} = n[F(X_{n,2}) - F(X_{n,1})] .$$

Then, by (2.1), (2.2), (2.7), (3.1) and (3.4) and proceeding as in the proof of Theorem 3.1 of Sen (1961), we obtain that

$$(3.5) \qquad E[Z_n - 2Y_{n(1)}^{a_m} + (Y_{n(1)} + Y_{n(2)})^{a_m}]^2 \to 0 \quad \text{as } n \to \infty .$$

and hence, by the Chebychev inequality, we have

$$(3.6)\quad \Lambda_m^*(x) = \lim_{n\to\infty} P\left\{2Y_{n(1)}^{a_m} - (Y_{n(1)} + Y_{n(2)})^{a_m} \le x\right\}, \quad \forall -\infty < x < \infty .$$

We may recall that $Y_{n(1)}$ and $Y_{n(2)}$ ar asymptotically independently distributed according to a common simple exponential law and they are non-negative rv's. For $x \le 0$, $\left[2Y_{n(1)}^{a_m} - (Y_{n(1)} + Y_{n(2)})^{a_m} \le x\right] \Longleftrightarrow \left[Y_{n(2)} \ge (2Y_{n(1)}^{a_m} - x)^{m+1} - Y_{n(1)}\right]$ and the first equation in (3.3) follows directly by finding the conditional probability given $Y_{n(1)}$ and then itegrating it out over $Y_{n(1)}$. For $x > 0$, if $Y_{n(1)} \le x^{m+1}$, then $2Y_{n(1)}^{a_m} - (Y_{n(1)} + Y_{n(2)})^{a_m} \le Y_{n(1)}^{a_m} \le x$, while for $Y_{n(1)} > x^{m+1}$, as before we need $Y_{n(2)} \ge (2Y_{n(1)}^{a_m} - x)^{m+1} - Y_{n(1)}$, and hence, the last equation in (3.3) follows on parallel lines. Q.E.D.

For $m = 0$ (i.e., $a_m = 1$), Λ_0 in (2.8) is the simple exponential while Λ_0^* in (3.3) is the double exponential df. For $m \ge 0$, Λ_m and Λ_m^* are not the same df.

<u>Theorem 2</u>. *For a terminal contact of order* $m(\ge 0)$,

$$(3.7)\quad \lim_{n\to\infty}\left\{E\left[b_{n,m}^2(\theta_n^* - \theta)^2\right]\right\} = \left\{1 - \frac{2a_m(1-a_m)}{1+a_m}\right\}\left[\lim_{n\to\infty}\left\{E\left[b_{n,m}^2(\hat{\theta}_n - \theta)^2\right]\right\}\right]$$

$$= \left(2a_m\Gamma 2a_m\{1 - 2a_m(1-a_m)/(1+a_m)\}\right) .$$

<u>Proof</u>. Since $\hat{\theta}_n = X_{n,1}$, by an appeal to Theorem 3.1 of Sen (1961), we get that

$$(3.8)\quad b_{n,m}^2 E(\hat{\theta}_n - \theta)^2 \to \Gamma 1+2a_m = 2a_m \Gamma 2a_m > 0 .$$

Hence, to prove (3.7), by (3.5), it suffices to show that as $n \to \infty$,

$$(3.9)\qquad E\left\{2Y_{n(1)}^{a_m} - (Y_{n(1)}+Y_{n(2)})^{a_m}\right\}^2 \to 2a_m\Gamma(2a_m)\{1 - 2a_m(1-a_m)/(1+a_m)\}\ .$$

Towards this, we may note that $E\left[Y_{n(1)}^{2a_m}\right] \to \Gamma(1+2a_m) = 2a_m\Gamma(2a_m)$, $E(Y_{n(1)}+Y_{n(2)})^{2a_n}$ $\to \Gamma(2+2a_m) = 2a_m(1+2a_m)\Gamma(2a_m)$ while $E\left\{Y_{n(1)}^{a_m}(Y_{n(1)}+Y_{n(2)})^{a_m}\right\} = E\left\{\left[E(Y_{n(1)}^{a_m}|Y_{n(1)}+Y_{n(2)})\right](Y_{n(1)}+Y_{n(2)})^{a_m}\right\} \sim E\left\{(Y_{n(1)}+Y_{n(2)})^{2a_m+1}/(a_m+1)\right\} \to \left\{\Gamma(2+2a_m)\right\}/(1+a_m) =$ $\left\{2a_m(1+2a_m)\Gamma(2a_m)\right\}/(a_m+1)$, and hence, (3.9) follows by some standard steps. Q.E.D.

For $m = 0$ (i.e., $a_m = 1$), the second factor on the rhs of (3.7) is equal to 1, so that both $\hat{\theta}_n$ and θ_n^* have the same asymptotic variance, though their df's are not the same. For $m \geq 1$ (i.e., $a_m \leq \frac{1}{2}$), $2a_m(1-a_m)/(1+a_m) > 0$ and is bounded from above by 1/3. Thus, from (3.2) and (3.7) we have that jackknifing reduces both the asymptotic bias and the asymptotic mean square to a fractional extent. This characteristic is different from the regular case where there is a complete reduction of asymptotic bias but no reduction of the asymptotic mean square.

From (2.6), (2.7) and (3.4), it follows that for a terminal contact of order $m(\geq 0)$,

$$(3.10)\qquad \left|n^{-1}b_{n,m}v_n - \left\{(Y_{n(1)}+Y_{n(2)})^{a_m} - Y_{n(1)}^{a_m}\right\}\right| \xrightarrow{p} 0\ ,\quad \text{as } n \to \infty\ .$$

Since $(Y_{n(1)}+Y_{n(2)})^{a_m} - Y_{n(1)}^{a_m} \to \left\{(Y_1+Y_2)^{a_m} - Y_1^{a_m}\right\}$, where Y_1 and Y_2 are i.i.d.r.v. having the simple exponential df on $[0,\infty)$, $n^{-1}b_{n,m}$ either converges to a positive constant (when $m = 0$) or goes to 0 (when $m \geq 1$), it follows that either (for $m = 0$) v_n has a non-degenerate asymptotic df

or (for $m \geq 1$) it goes to $+\infty$, in probability as $n \to \infty$. This characteristic is also different from the regular case where $v_n \xrightarrow{p} a$ constant, as $n \to \infty$. Nevertheless, for the studentized form, we have for a terminal contact of order $m (\geq 0)$,

$$T_n = n(\theta_n^* - \theta)/v_n = b_{n,m}(X_{n,1} - \theta)/b_{n,m}(X_{n,2} - X_{n,1}) - (n-1)/n$$

$$+ o_p(1) \xrightarrow{\mathcal{D}} Y_1^{a_m} / \left\{ (Y_1 + Y_2)^{a_m} - Y_1^{a_m} \right\} - 1 , \tag{3.11}$$

so that noting that $Y^* = Y_2/Y_1$ has the Fisher's variance-ratio distribution with degrees of freedom (2,2), we have from (3.11) that

$$[1 + (1 + T_n)^{-1}]^{m+1} - 1 \xrightarrow{\mathcal{D}} Y^* = Y_2/Y_1 . \tag{3.12}$$

For $m = 0$, we have a simplified form

$$T_n + 1 \xrightarrow{\mathcal{D}} Y_1/Y_2 \overset{\mathcal{D}}{=} Y^* . \tag{3.13}$$

Both (3.12) and (3.13) have important statistical applications.

4. A MODIFICATION OF θ_n^*

We have observed in (3.2) that for $m \geq 1$, $b_{n,m} E(\theta_n^* - \theta)$ does not converge to 0 as $n \to \infty$. Let C_n be the sigma-field generated by $X_{n,1}, \ldots, X_{n,n}$ and by X_{n+j}, $j \geq 1$ (so that C_n is non-increasing in n). Then, in the regular case, [cf. (2.11) of Sen (1977)], we have

$$\theta_n^* - \hat{\theta}_n = (n-1) E\{(\hat{\theta}_n - \hat{\theta}_{n-1}) | C_n\} . \tag{4.1}$$

In our case, for $m \geq 1$, $nb_{n,m}E(\hat{\theta}_n - \hat{\theta}_{n-1}) = -a_m\sqrt{1+a_m} + o(1)$, where as $b_{n,m}E(\hat{\theta}_n - \theta) = \sqrt{1+a_m} + o(1)$, and thereby, we get the resulting bias in (3.2). To eliminate the, we may consider the modified estimator

$$\theta^{**}_{n,m} = \hat{\theta}_n + \frac{1}{a_m} E\{(\hat{\theta}_n - \hat{\theta}_{n-1})|C_n\} \tag{4.2}$$

$$= X_{n,1} - (m+1)n^{-1}(n-1)(X_{n,2} - X_{n,1}) .$$

In that case, we have

$$b_{n,m}E(\theta^{**}_{n,m} - \theta) \to 0 \quad \text{as} \quad n \to \infty . \tag{4.3}$$

Also, following the same line as in the proof of Theorem 1, we obtain that

$$\Lambda^{**}_m(x) = \lim_{n\to\infty} P\{b_{n,m}(\theta^{**}_{n,m} - \theta) \leq x\}$$

$$= \begin{cases} \int_0^\infty \exp\left\{-\left[\{(m+2)y^{a_m} - x\}/(m+1)\right]^{m+1}\right\}dy, & -\infty < x \leq 0 , \\ 1-\exp\{-x^{m+1}\} + \int_{x^{m+1}}^\infty \exp\left\{-\left[\{(m+2)y^{a_m} - x\}/(m+1)\right]^{m+1}\right\}dy, & 0 < x < \infty . \end{cases} \tag{4.4}$$

Also, following the line of proof of Theorem 2, we have

$$\lim_{n\to\infty} E\left\{b^2_{n,m}(\theta^{**}_{n,m} - \theta)^2\right\} = (2a_m\sqrt{2a_m})\left\{1 - \frac{2a_m}{1+a_m}(m+1)[(m+1)a_m - 1]\right\} = 2a_m\sqrt{2a_m}$$

$$= \lim_{n\to\infty} E\left\{b^2_{n,m}(\hat{\theta}_n - \theta)^2\right\} \geq \lim_{n\to\infty} E\left\{b^2_{n,m}(\theta^*_n - \theta)^2\right\} . \tag{4.5}$$

Thus, whereas $\theta^{**}_{n,m}$ eliminates bias to the desired extent, it fails to reduce the mean square. In this sense, it is similar to the case of θ^*_n in the regular case. [Though Λ^{**}_m and Λ_m are not the same.]

Finally, for the studentized case, in (3.11)-(3.13), the only changes we need to made is to replace T_n by $T_n + m$; the rest remains the same.

5. SOME REMARKS

We have so far considered the case of the lower end-point. The case of the upper end-point (if finite) follows on parallel lines. Secondly, in practical applications, when the form of F is not specified but the order of terminal contact is assumed to be known [viz., $m = 0$ when F is U-shaped or inverted J-shaped, etc.], the studentized form in (3.11)-(3.13) may most conveniently be used to provide a jackknife test for a null hypothesis $H_0: \theta = \theta_0$ (specified) or a confidence interval for the unknown θ. For a symmetric df with both end-points finite, jackknifing of the extreme mid-range (for estimating or testing for the location of the df) can be made — the jackknife estimator corresponding to the smallest and the largest order statistic are also asymptotically independent.

REFERENCES

[1] Arvesen, J.M. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* 40, 2076-2100.

[2] Schucany, W.R., Gray, H.L. and Owen, D.M. (1971). Bias reduction in estimation. *Jour. Amer. Statist. Assoc.* 66, 524-533.

[3] Sen, P.K. (1961). A note on the large sample behaviour of extreme sample values from distributions with finite end-points. *Calcutta Statist. Assoc. Bull.* 10, 106-115.

[4] Sen, P.K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. *Ann. Statist.* 5, 316-329.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 78-1271	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON JACKKNIFING IN ESTIMATING THE FINITE END-POINTS OF A DISTRIBUTION		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Pranab Kumar Sen		8. CONTRACT OR GRANT NUMBER(s) AFOSR 74-2736
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina Department of Biostatistics Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE 1978
		13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

bias, extreme values; jackknifing; mean square; order of terminal contact; studentized form; Tukey-estimator of variance.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Sample extreme values are biased estimators of the end-points of a distribution, and hence, jackknifing is useful. However, the properties of jackknifing in such a case differ considerably from those in the regular case. These are studied here. Along with modification of jackknifing, some applications are also considered.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)